

Vera Hildenbrandt, Annette Klosa

Lexikographische Prozesse bei Internetwörterbüchern

Einleitung

Während lexikographische Prozesse,¹ die zur Publikation gedruckter Wörterbücher führen, bereits seit einigen Jahrzehnten im Fokus der Wörterbuchforschung stehen² und die von Wiegand dafür unterschiedenen Phasen der Vorbereitung, der Datenbeschaffung, der Datenaufbereitung, der Datenauswertung und der Satz- und Druckvorbereitung (Wiegand et al. 2010, S. 8ff.) mittlerweile als etabliert betrachtet werden dürfen,³ steht die Diskussion und Beschreibung lexikographischer Prozesse von Internetwörterbüchern noch in den Anfängen. Zwar besteht kein Zweifel daran, dass sich lexikographische Prozesse bei der Publikation von Internetwörterbüchern anders gestalten als bei Printwörterbüchern, doch die Fragen, inwiefern sie dies tun, welchen Einfluss die neuen Möglichkeiten der Datengewinnung aus elektronischen Textkorpora auf die Prozesse haben, wie Bearbeitungsteilwortschätze auszuwählen sind, wie verschiedene Fassungen zu versionieren und zu archivieren sind und wie sich schließlich die Änderungen der lexikographischen Prozesse auf die Nutzer auswirken, ob und wie die Nutzer in diese Prozesse einbezogen werden können, sind noch nicht ausführlich beantwortet. Diese und andere Fragen waren daher Gegenstand des vierten Arbeitstreffens des wissenschaftlichen Netzwerks „Internetlexikografie“, das am 22. und 23. November 2012 an der Universität Trier stattfand und vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften/Trier Center for Digital Humanities organisiert wurde.

Im Rahmen des Arbeitstreffens wurden zunächst kurze und vertiefte Einblicke in die lexikographischen Prozesse verschiedener Wörterbücher gegeben. Mit dem Trierer Wörterbuchnetz (www.woerterbuchnetz.de) wurde eine lexikographische Online-Publikation präsentiert, die ausschließlich aus retrodigitalisierten, überwiegend historisch ausgerichteten Wörterbüchern besteht. Auch zu den lexikalischen Ressourcen des Digitalen Wörterbuchs der deutschen Sprache (www.dwds.de) gehören retrodigitalisierte Wörterbücher. Darüber hinaus aber entwickelt das DWDS ein korpusbasiertes Digitales Lexikalisches System, das die Wörter des deutschen Wortschatzes in ihren Eigenschaften beschreibt.⁴ Das Mittelhochdeutsche Wörterbuch (www.mhdwb-online.de) steht zwischen gedruckt und online erscheinendem Nachschlagewerk; alle Phasen seiner Erstellung werden computerunterstützt durchgeführt, konzeptionell führend ist allerdings das gedruckte Wörterbuch, die nahezu zeitgleich erscheinende Internetpublikation wird davon abgeleitet. Auch Duden online (www.duden.de, vgl. den Beitrag von Melina Alexa, Anja Konopka und Lutz Wind) entsteht aus derselben Datenbasis wie die gedruckten Wörterbücher des Dudenverlags, allerdings gibt es medienspezifische Unterschiede im Her-

¹ Bewusst sei hier der Plural verwendet, da es – dies hat sich nicht zuletzt in den Diskussionen des Arbeitstreffens gezeigt – den lexikographischen Prozess schlechthin nicht gibt, da dessen Ausgestaltung von verschiedenen Faktoren – u.a. etwa von Typ und Entstehungsweise des Wörterbuchs – abhängt.

² Einen knappen Überblick über den Stand der Forschung gibt Klosa (2013, S. 519).

³ Vgl. etwa die Beiträge in diesem Arbeitsbericht, die sich nahezu ausnahmslos auf Wiegand beziehen.

⁴ Vgl. dazu Geyken (2014).

stellungsprozess der online und der im Print erscheinenden Ressourcen. Mit dem Algemeen Nederlands Woordenboek (<http://anw.inl.nl/search>, vgl. den Beitrag von Carole Tiberius und Tanneke Schoonheim) und elexiko (www.owid.de/wb/elexiko/start.html, vgl. den Beitrag von Annette Klosa) entstehen von vornherein digital konzipierte Online-Wörterbücher der niederländischen respektive deutschen Gegenwartssprache, mit dem Digitalen Familiennamenwörterbuch Deutschlands (vgl. den Beitrag von Sandra Denzer und Franziska Horn) ein „digital born“-Spezialwörterbuch deutscher Familiennamen. Ebenfalls ausschließlich für die Online-Publikation entwickelt ist das allgemeine, mehrsprachige Wörterbuch Wiktionary (<http://de.wiktionary.org/wiki/Wiktionary:Hauptseite>, vgl. den Beitrag von Christian M. Meyer und Iryna Gurevych), das im Gegensatz zu den bereits angeführten Wörterbüchern allerdings nicht redaktionell, sondern gemeinschaftlich erarbeitet wird.

Ob nun retrodigitalisiert, für den Druck konzipiert und dann online publiziert oder ausschließlich digital entworfen und im Internet veröffentlicht, ob abgeschlossenes oder Ausbauwörterbuch, ob redaktionell als Verlagsprodukt, von einer wissenschaftlichen Einrichtung erarbeitetes oder kollaborativ erstelltes Werk – die Einblicke verdeutlichten, dass sich die lexikographischen Prozesse dieser Internetwörterbücher oder -informationssysteme von denen ausschließlich gedruckt erscheinender Wörterbücher unterscheiden. Zugleich aber demonstrierten sie, dass die lexikographischen Prozesse abhängig sind von der Art des Internetwörterbuchs, dass also etwa der eines gemeinschaftlich erstellten Wörterbuchs anders ist als der eines redaktionell betreuten. Und schließlich zeigten sie die Wörterbuchredakteure auf der Suche nach einer adäquaten Beschreibung der lexikographischen Prozesse. Den praktischen Einblicken folgte ein theoretischer Beitrag, der auf methodologischer und metalexikographischer Ebene die Notwendigkeit einer allgemeinen Lexikographietheorie diskutierte, die als Basis dienen kann für die Definition verschiedener Arten von lexikographischen Prozessen von Print- und Internetwörterbüchern (vgl. den Aufsatz von Rufus Gouws).

Die Auseinandersetzung mit dem lexikographischen Prozess wurde fortgesetzt in drei Arbeitsgruppen, die sich mit Auswahlkriterien, Umsetzung und Problemen von Bearbeitungsteilwortschätzen, mit Archivierung und Versionierung und mit dem korpusbasierten Vorgehen bei der Erweiterung bestehender lexikographischer Ressourcen beschäftigten. Auf die eingangs formulierten Fragen wurden in den Diskussionsrunden und Arbeitsgruppen die folgenden Ergebnisse gefunden bzw. weiterführende Fragen identifiziert.⁵

1. Inwiefern gestalten sich lexikographische Prozesse bei der Publikation von Internetwörterbüchern anders als bei Printwörterbüchern?

Dass beim Versuch der Beschreibung lexikographischer Prozesse von Internetwörterbüchern immer wieder Bezug genommen wird auf das für die Entstehung gedruckter Wörterbücher entwickelte 5-Phasen-Modell Wiegands,⁶ verdeutlicht, wie sehr die Internetlexikographie sich hier noch in einer Zeit des Umbruchs befindet. Besonders redaktionell erstellte Internetwörterbücher berufen sich zunächst auf diese Phasierung, erweitern sie allerdings um eine Phase

⁵ Die im Folgenden gegebene Zusammenfassung der Diskussion basiert auf den Protokollen der Hilfskräfte des Trierer Kompetenzzentrums: Sandra Behrend, Semjon Borchert, Katharina Dietz, Matthias Schneider, Aline Stang und René Tobner sei an dieser Stelle herzlich gedankt.

⁶ Vgl. die Beiträge von Annette Klosa, Christian Meyer/Iryna Gurevych, Carole Tiberius/Tanneke Schoonheim und Rufus Gouws.

der Digitalisierung⁷ oder Computerisierung,⁸ in der die nötige Software bereitgestellt wird und Analysewerkzeuge zum Einsatz kommen. Die konkrete Umsetzung der einzelnen Phasen allerdings unterscheidet sich in Abhängigkeit vom entstehenden Wörterbuchtyp zum Teil fundamental von denen gedruckter Wörterbücher. Dies gilt in besonderer Weise für die letzte Phase in diesem Prozess, die Publikationsphase, die medienspezifisch anders abläuft, weniger auf ein abgeschlossenes als vielmehr auf ein offenes und dynamisches Produkt abzielt, das in regelmäßigen Abständen aktualisiert werden kann oder sich – wie Wiktionary – in einem beständigen Aktualisierungsvorgang befindet. Auf kollaborativ erarbeitete Wörterbücher scheinen denn auch die aus Printwörterbüchern bekannten Phasen nur in sehr begrenztem Maße anwendbar zu sein. Für Wiktionary wird daher ein neues, auf eine Vorbereitungs- und eine Bearbeitungsphase reduziertes Modell des lexikographischen Prozesses vorgeschlagen.⁹ Dabei werden diese Phasen – das wiederum ist allen im Rahmen des Arbeitstreffens vorgestellten Wörterbüchern gemeinsam – iterativ durchlaufen. Charakteristisch für lexikographische Prozesse bei Internetwörterbüchern ist also ein zirkulärer, nicht-linearer Verlauf, an dem im Vergleich zu Printwörterbüchern nicht überwiegend Lexikographen, sondern auch Computerlinguisten und Informatiker beteiligt sind und in den auch die potenziellen Nutzer verstärkt einbezogen werden können und müssen, sollten sie denn nicht ohnehin wie im Falle von Wiktionary gleichzeitig auch Wörterbuchautoren sein. Voraussetzungen für die produktive Gestaltung des lexikographischen Prozesses und das effiziente Einbeziehen des Nutzers sind ein vollständiges Durch- und immer wieder Überdenken dieses Prozesses und seiner Transparenz.

2. Wie sind Bearbeitungsteilwortschätze auszuwählen?

Die Nutzung von Internetwörterbüchern kann bereits gewährleistet sein, bevor sie vollständig bearbeitet sind. Dies wird u.a. durch die Arbeit in und die Publikation von Bearbeitungsteilwortschätzen ermöglicht, deren Auswahl in der Vorbereitungsphase des lexikographischen Prozesses mitgeplant werden sollte. Für die Arbeit in Bearbeitungsteilwortschätzen sprechen nicht nur mediale, sondern auch redaktionelle Gründe. Sie dient der Rationalisierung der lexikographischen Arbeit, erlaubt die systematische und umfassende Beschreibung vergleichender Wortschatzeinheiten und die Nutzung spezieller Expertise für bestimmte Wortschatzbereiche wie zum Beispiel etymologische Angaben oder Fachwortschätze. Allerdings bringt die Arbeit in Bearbeitungsteilwortschätzen auch Probleme mit sich. So entstehen etwa Reibungsverluste, wenn die auf Alphabetstrecken oder nach Wortarten verteilte Bearbeitung die konsistente Vernetzung der Stichwörter erschwert oder die Arbeit auf verschiedene Autorengruppen verteilt wird. Darüber hinaus ist es nicht ausgeschlossen, dass die Auswahl und Bearbeitung von Bearbeitungsteilwortschätzen an den Interessen der Nutzer vorbeigeht, da sie von projektinternen Gegebenheiten gesteuert wird. Soll die Bearbeitung in Bearbeitungsteilwortschätzen gelingen, müssen Lexikographen regelmäßig ihre Methoden überdenken und gemeinsam mit den Benutzern immer wieder eine reflexive Phase einschieben. Darüber hinaus sollte ein verständliches und erweiterbares Instruktionsbuch für eine gute Planung, Disziplin, Flexibilität, Kontrolle, Motivation und Transparenz sorgen.

⁷ Vgl. den Beitrag von Carole Tiberius und Tanneke Schoonheim.

⁸ Vgl. den Beitrag von Annette Klosa.

⁹ Vgl. den Beitrag von Christian Meyer und Iryna Gurevych.

3. Welchen Einfluss haben die neuen Möglichkeiten der Datengewinnung aus elektronischen Textkorpora auf die Prozesse?

Transparenz ist auch gefordert beim korpusbasierten Vorgehen in der lexikographischen Arbeit. Zur Beantwortung der Frage, welchen Einfluss die neuen Möglichkeiten der Datengewinnung aus elektronischen Korpora auf die lexikographischen Prozesse haben, wurden verschiedene Thesen aufgestellt:

These 1: Der Nutzen des Korpus und der Korpusrecherche hängt vom konkreten lexikographischen Ziel ab.

These 2: Ein Korpus ist nur so gut wie: a) die Abfragesprache, b) die Metadaten, c) die Größe eines sichtbaren Textausschnitts.

These 3: Gute Werkzeuge filtern und sortieren die Daten und machen sie handlich für die lexikographische Arbeit. Sie „übersehen“ aber alles, was selten ist.

These 4: „There is no data like more data.“

These 5: Das, was ins Wörterbuch aufgenommen wird, entspricht nicht dem, was im Korpus steht; die sprachliche Wirklichkeit steht also der lexikographischen Bearbeitung entgegen.

Aus diesen Thesen wurden für die Arbeit mit Korpusdaten im Wörterbuchprozess einige Forderungen abgeleitet: Korpustechnologen sollten nicht nur Statistiken und vorgefilterte Daten anbieten, sondern auch Data Mining-Funktionen wie z.B. Verteilungsdiagramme bereitstellen. Auch sollten sie automatisierte Mechanismen zur Kontrolle der sich verändernden Korpuszusammensetzung zur Verfügung stellen. Metadaten der Korpora sollten im Dialog mit Lexikographen erstellt und abfragbar gemacht werden. Lexikographen müssen im Umgang mit den Korpusdaten die Möglichkeit erhalten, die Daten zu annotieren. Dazu ist es jedoch zum einen erforderlich, dass sie in der Lage sind, über das Korpus hinauszublicken und zum anderen die statistisch vorgefilterten Daten kritisch analysieren zu können. Nur auf diese Weise können sie die Nutzer für die unterschiedliche Qualität der Datenarten sensibilisieren und geeignete Recherchetechniken in Wörterbuch- und Korpusdaten vermitteln. Den Nutzern sollte es in jedem Fall ermöglicht werden, die Sprachwirklichkeit weitgehend ungefiltert rezipieren zu können, d.h. neben dem Wörterbuch sollte das Korpus selbst in Gänze als eigenständige Publikation für verschiedene Anwendungen zur Verfügung stehen.

4. Wie sind verschiedene Fassungen eines Internetwörterbuchs zu versionieren und zu archivieren?

Auch Versionierung und Archivierung sind Teil der lexikographischen Prozesse und müssen bereits bei der Projektplanung bedacht werden. In Bezug auf die Versionierung besteht grundsätzlich kein Zweifel daran, dass die Daten permanent im jeweils aktuellen Zustand präsentiert werden und dass transparente Versionsmetadaten zur Verfügung stehen sollten, sodass für den Nutzer jederzeit ersichtlich ist, in welcher Wörterbuchversion er nachschlägt. Eine Timeline könnte hier gegebenenfalls für einen schnellen Überblick sorgen. Auch in Bezug auf die Archivierung gibt es eindeutige Forderungen. Die Wörterbuchdaten und -metadaten soll-

ten möglichst unabhängig von den technischen Bedingungen und Gegebenheiten in gut dokumentierten, offenen, standardisierten Formaten vorgehalten werden. Dabei sollten nicht nur Text-, sondern auch andere Medienformate wie beispielsweise Bilder und Karten bedacht werden. Die Daten sollten persistent zitier- und über Persistent Identifier adressierbar sein, wobei frühzeitig geklärt werden sollte, wie feingranular die Persistent Identifier sein können und sollen.

Aber auch offene Fragen und Probleme wurden in der Diskussion identifiziert. Ungeklärt ist etwa, ob Nutzer ein Interesse an früheren Versionen haben. Und auch die langfristige Vorhaltung von Infrastrukturen ist noch immer eine Herausforderung. Zwar ist die Notwendigkeit von Ressourcenkuration und -pflege, von Standardisierung und Archivierung zwischenzeitlich im Denken digital arbeitender Wissenschaftler verankert und wird auch von den Forschungsförderern gefordert,¹⁰ doch stehen für diesen wichtigen Teil digitaler wissenschaftlicher Tätigkeiten nach wie vor keine ausreichenden Förderformate zur Verfügung. Als andauernde Aufgabe ist in diesem Kontext die Ausbildung eines öffentlichen Bewusstseins für die neue digitale Kultur zu sehen.

5. Wie wirken sich Änderungen lexikographischer Prozesse auf Nutzer aus? Wie können Nutzer in diese Prozesse einbezogen werden?

Die potenziellen Nutzer von Internetwörterbüchern gehören zu dieser Öffentlichkeit und sollten verstärkt und von Beginn an in die lexikographischen Prozesse einbezogen werden. Der Umfang des Einbezogenwerdens und Einbezogenwerdenkönnens sowie des Nutzereinflusses richtet sich dabei natürlich nach dem Typ des Wörterbuchs und wird bei einem kollaborativ entstehenden Crowd-Sourcing-Wörterbuch wie Wiktionary größer sein als bei einem redaktionell erarbeiteten Wörterbuch wie *eleXiko*. Einigkeit besteht jedoch darin, dass Nutzer bereits in der Planungs- oder Vorbereitungsphase des lexikographischen Prozesses eine wesentliche Rolle spielen sollten. Nach Möglichkeit sollten die verschiedenen Benutzergruppen identifiziert und die Funktionen des Wörterbuchs an ihren Bedürfnissen ausgerichtet, also unterschiedliche, unter Umständen auch interaktiv einzurichtende Nutzerprofile vorgesehen werden. Auch sollten die Nutzer jederzeit Rückmeldung geben können, was letztlich auch Auswirkungen auf die weitere lexikographische Arbeit haben kann.¹¹ Wichtig sind also zu jedem Zeitpunkt des lexikographischen Prozesses die Kommunikation mit den Nutzern, Offenheit und Transparenz.

Wie allerdings diese Transparenz geschaffen, für die Nutzer dargestellt werden kann, ist noch eine ebenso offene Frage wie die zukünftigen Nutzerbedürfnisse ungeklärt sind. Die Notwendigkeit empirischer, auch auf eine mittel- bis langfristige Perspektive gerichteter Nutzerstudien ist also offenkundig. Zugleich aber sollten Lexikographen, Wissenschaftler und versierte Wörterbuchnutzer angehalten werden, Nutzungsmöglichkeiten zu offerieren und zu erklären, um so Potenziale aufzuzeigen und Interesse und Bedarf zu wecken.

Insgesamt führte das Arbeitstreffen in aller Deutlichkeit vor Augen, wie sehr lexikographische Prozesse (in Abhängigkeit vom Wörterbuchtyp) im Wandel begriffen sind, wie sehr sich die zur Bewältigung der Prozesse notwendigen Kompetenzen und damit Berufsgruppen ausweiten, wie sehr die Nutzer von Beginn an einbezogen werden sollten und wie wichtig Transpa-

¹⁰ Vgl. etwa die DFG-Praxisregeln „Digitalisierung“ (Deutsche Forschungsgemeinschaft (Hg.) 2013).

¹¹ Zur Rolle des *simultaneous feedback* im lexikographischen Prozess vgl. den Beitrag von Rufus Gouws.

renz und Kommunikation in allen Prozessphasen sind. Es zeigte aber auch, wie viele offene Fragen es noch gibt und wie wichtig eine theoretische Reflexion der lexikographischen Prozesse bei Internetwörterbüchern ist.

Literatur

- Deutsche Forschungsgemeinschaft (Hg.) (2013): DFG-Praxisregeln „Digitalisierung“. DFG Vordruck 12.151-02/13. Bonn: Deutsche Forschungsgemeinschaft. www.dfg.de/formulare/12_151/12_151_de.pdf (Stand: 30.7.2015).
- Geyken, Alexander (2014): Methoden bei der Wörterbuchplanung in Zeiten der Internetlexikographie. In: *Lexicographica* 30, S. 77-112.
- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus Hjalmar/Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herbert Ernst (Hg.): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. (= *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)* 5.4). Berlin/Boston: de Gruyter, S. 517-524.
- Wiegand, Herbert Ernst/Beißwenger, Michael/Gouws, Rufus H./Kammerer, Matthias/Storrer, Angelika/Wolski, Werner (Hg.) (2010): *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*. Bd. 1: A-C. Berlin/New York: de Gruyter.